UCRL-53656

Guidelines for the Integration of Audio Cues into Computer User Interfaces

D. A. Sumikawa

(M.S. Thesis)

June, 1985

Lawrence Livermore National Laboratory

DISCLAIMER

Work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

UCRL-53656
Distribution Category UC-32

Guidelines for the Integration of Audio Cues into Computer User Interfaces

D. A. Sumikawa

(M.S. Thesis)

Manuscript date: June, 1985

LAWRENCE LIVERMORE NATIONAL LABORATORY
University of California • Livermore, California • 94550

Available from: National Technical Information Service • U.S. Department of Commerce
5285 Port Royal Road • Springfield, VA 22161 • $10.00 per copy • (Microfiche $4.50)

Guidelines for the Integration of Audio Cues into Computer User Interfaces

By

DENISE ANN SUMIKAWA
B.A. (University of Colorado, Boulder) 1979

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE
in

Computing Science

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Meera M Blattner

Ken I. Joy

Charles U. Martel

Committee in Charge

Deposited in the University Library ________ Date ________ Librarian

Abstract

Throughout the history of computers, vision has been the main channel through which information is conveyed to the computer user. As the complexities of man-machine interactions increase, more and more information must be transferred from the computer to the user and then successfully interpreted by the user. A logical next step in the evolution of the computer-user interface is the incorporation of sound and thereby using the sense of "hearing' in the computer experience. This allows our visual and auditory capabilities to work naturally together in unison leading to more effective and efficient interpretation of all information received by the user from the computer. This thesis presents an initial set of guidelines to assist interface developers in designing an effective sight and sound user interface. This study is a synthesis of various aspects of sound, human communication, computer-user interfaces, and psychoacoustics. We introduce the notion of an earcon. Earcons are audio cues used in the computer-user interface to provide information and feedback to the user about some computer object, operation, or interaction. A possible construction technique for earcons, the use of earcons in the interface, how earcons are learned and remembered, and the affects of earcons on their users are investigated. This study takes the point of view that earcons are a language and human/computer communication issue and are therefore analyzed according to the three dimensions of linguistics; syntactics, semantics, and pragmatics.

Table of Contents

PART I.

OVERVIEW

CHAPTER 1

Introduction

Throughout the history of computers, vision has been the main channel through which information is conveyed to the computer user. Data is displayed visually via the computer terminal or some off-line printer. As the complexities of the man-machine interactions continue to increase, more and more information must be transferred from the computer to the user and then successfully interpreted by the user. Increasing the number of visually displayed items on a computer screen, however, increases time and errors in visual searching, counting, and the noticing of display changes [Ram79]. A high density of visually displayed information also cause a "cognitive overload" which effects a user's decision making performance. Fortunately, vision is only one of the five human sensory channels which are available for information transference. People by nature are not merely visual beings. We use all of our senses to interpret the information we receive from our environment.

A logical next step in the evolution of the computer/user interface is the incorporation of sound and thereby using the sense of "hearing" into the computer experience. This opens up another vital communication and information channel between man and machine. It can provide relief to the often overloaded visual link. Communication is defined as "the exchange of messages or information by speech, signals, or writing". This implies the use of both vision and audition. Man communicates with the computer during interactive sessions. Therefore, sight and sound need to be integrated to allow a natural man-machine conversation. There is a need to develop interactive computer systems that are more compatible with the abilities of the potential users of that system. When the communication interface between man and machine closely resembles man's natural mode of communication, people will make better use of the computer. It allows our visual and auditory capabilities to work naturally together in unison to lead to a more effective and efficient interpretation of all information received by the user from the computer.

The idea of a sight and sound communication interface is still in its infancy. Various work has been done on incorporating sound in the computer. These include music, speech synthesis, speech recognition, and the beeps, buzzes, and bells of video games. This is a good start at recognizing the importance of hearing and sound in computer systems.

The goal of an effective sight/sound interface can only be realized by a thorough understanding of how man uses sound to communicate in his natural environment. By investigating various aspects of sound, human communication, and computer interfaces, this thesis presents useful information and an initial set of design guidelines to assist interface developers in designing an effective sight/sound interface between the computer and its intended user population.

The specific aspect of sound and the computer/user interface that this study addresses is the addition of audio cues to various components of the user interface. These are sounds that range from singular to complex, soft to loud, subtle to unavoidable, intermittent to continuous, and background to foreground. Components of the user interface include editors, system messages, menus, windows, errors, and prompts to name just a few. As an example, consider the possibility of adding a soft subtle background sound to indicate your current editing mode of either deleting or inserting. Or a certain tone to indicate that mail has just arrived on the system for you. Or perhaps a louder sound to accompany a system message that is alerting you to the fact that the computer will be going down in five minutes. Often times the visual equivalent of these situations escapes our attention and numerous mistakes are made and valuable time is lost. Sounds are especially useful when an interruption of the visual display currently on the screen is undesirable.

The sounds or tones mentioned above are all examples of auditory cues. To coin a new word, we are introducing the term *earcon* which can be considered a synonym of audio cue. Earcons are the audio analogy to the visual and graphical ICONS ("eye"cons) that are so prevalent. By definition, *earcons are audio cues that are used in the computer/user interface to provide information to the user about some computer object, operation, or interaction.*

A word should be said on the exclusion of voice output from this study. Although voice is important, it introduces a different dimension into the study of sound in the computer/user interface. Voice adds information beyond pure sound and was not included because we had to limit the scope of this study. Also, details concerning physical implementation of earcons are not addressed.

1.1. Methodology

At the present time, information concerning the integration of sound in a man/machine interface, other than speech or music, is very scarce. There is a definite lack of information specifying how an interface designer should proceed to incorporate sound into a computer/user interface.

The design guidelines presented here are an initial attempt to provide valuable information to begin filling this void. An extensive search on various disciplines was conducted to gather pertinent data which can be compiled and integrated to form a set of useful guidelines concerning the addition of sound to the computer/user interface. The disciplines analyzed include computer science, psychology, psychoacoustics, information theory, communication theory, radio advertising, music, human factors, and the theory of sound. Each of these areas contributed in a distinct way to the formation of the design guidelines. A few informational conflicts were encountered between the different sources analyzed. However, it is not the purpose of this thesis to resolve such conflicts but only to indicate where they did occur. Every attempt was made to insure that the guidelines were formulated in a systematic way, basing them on established results from the different disciplines instead of using opinions and personal preferences.

The present set of guidelines are by no means complete and are not meant to be presented as standards. They are merely plausible suggestions, based on various results that the author believes to be useful. Since many of the guidelines are general principles or suggestions, human factors experimentation is desirable before a specific implementation is made. The purpose of presenting this information is to propose a basic set of design guidelines and to present some initial ideas that, hopefully, will stimulate interest in this area and motivate future research.

The guidelines stated in the following chapters are presented in italics and are marked as being one of two types; either principle (P) or measurable (M). Principle guidelines present general information while the measurable guidelines state more specific information that can be measured through empirical experimentation.

1.2. Organizational Structure

Earcons communicate information between man and machine through audio signs and signals. Being a nonverbal form of communication, it is appropriate to analyze the various aspects and problems of "earcons" according to the three dimensions of semiotics - the science of signs.

Semiotics is a general philosophical theory of signs and symbols that deals with their function in both artificially and naturally constructed languages [Web77]. It calls attention to the three dimensions of communication; syntactics, semantics, and pragmatics. Syntactics is concerned with the formal relations between signs in abstraction from their significance and their interpretors. Semantics deals with the relation between signs and what they refer to. Pragmatics is concerned with the relation between signs and their users. In each of these dimensions, signs communicate with human observers.

The organizational structure of this thesis is modeled after the dimensions of semiotics. A fourth area, the lexical aspects of earcons, is also included to investigate the basic characteristics of sound. Part II examines the building blocks of audio cues (lexical), Part III investigates how these building blocks are arranged to form earcons (syntactics), Part IV studies the meaning of earcons (semantics), and Part V explores the human element of audio messages (pragmatics).

CHAPTER 2

Earcons

2.1. Description

Earcons were defined in Chapter 1 as being audio cues that are used in the computer/user interface to provide information to the user about some computer object, operation, or interaction. The basic purpose of an earcon is to provide an alternate channel through which information, or messages, can be communicated from the computer to the user. Any message in a typical programming environment can be represented by sound. Therefore, earcons can be considered as sound messages in which information is conveyed. Audio messages can be just as effective, often more effective, than their graphical counterparts. A person's field of hearing is 360 degrees whereas their field of vision is less than 180 degrees. Textual messages often escapes a user's field of vision but an audio message can reach a user even if they are not sitting directly in front of a terminal.

Audio messages can be used to either replace or accompany textual messages. Many messages can be conveyed quite well through sound only, thereby freeing visual space on the screen. Other messages may need to be emphasized and an audio cue accompanying the textual message is effective in this respect.

Each earcon represents one computer entity. Here, a computer entity is defined to be an object, operation, or interaction. Examples of computer objects are files, menus, and prompts. Editing, compiling, and executing are examples of operations. An example of an interaction between an object and an operation is editing a file. There are countless such computer entity in a typical interactive session. The user learns which audio cue represents which computer entity through a learning process, much like we learn which English word represents which object when we are young. Chapter 8 addresses this in more detail.

Several types of earcons exist, each dependent on the type of computer entity that the earcon is depicting. Certain events are urgent and must be represented by an urgent sounding audio cue. Urgent earcons are of utmost importance to the user and require their immediate attention. Action earcons are those requiring some type of user input. These call for a user's attention and therefore should be composed of tones capable of capturing their attention. Other audio cues don't require a user's attention but are there merely as an information supplement to provide additional feedback during an interactive session. No real damage is done if a user fails to notice this type of earcon. An example of this would be a soft, unobtrusive background audio cue indicating that new mail has just arrived.

Different earcons have different tonal compositions. The type of sound elements composing an audio cue is dependent on the type of entity that the particular earcon depicts. For example, an urgent system message could be represented by loud, repetitious tones. An editing mode, however, could be represented by a very soft, almost unnoticeable sound. Earcons differ in loudness, repetitions, and length. These variables are also dependent on what a particular audio cue is trying to represent to the user. Chapters 4, 5, 6, and 7 address the above aspects of audio cues and present guidelines and further information concerning their composition.

2.2. Benefits

Earcons, or audio cues, have many benefits which include the following:

- Provides a natural communication link between man and machine.
- Relieves an overloaded visual channel by presenting some messages in sound.
- Allows more information to be simultaneously transmitted to the user via sight and sound.
- Allows emphasis of graphical messages by accompanying them with an audio message.
- Provides an alternate method of capturing a user's attention.
- Relieves the user of having visual contact with the computer screen, thereby allowing the user more freedom of movement.

2.3. Examples

Several examples have already been given throughout this chapter. Additional examples are now provided to show the various ways audio cues can be successfully integrated into a computer/user interface and to demonstrate their versatility.

(1) Audio cues representing urgent system messages such as the system going down in a few minutes.

(2) Audio cues representing system messages pointing out that a user's files were destroyed during a disk crash.

(3) Earcons representing system prompts that require user input for continuation.

(4) Audio cues representing the completion of a program's compilation. Users often waste valuable time by failing to notice the visual message that compilation has finished. An audio announcement of such completion would be more effective in gaining a user's attention to this fact.

(5) An earcon immediately indicating misspelled words in user input. These don't have to be obnoxious beeps but could be some soft, subtle, background sound.

(6) An audio cue that is played when a user's code is successfully written to disk.

(7) An earcon to indicate an out of bounds variable when in a structured editor.

(8) Audio cues that denote fatal errors during program execution.

(9) Earcons that signify different types of error messages or accompany a textual error message for emphasis.

(10) Earcons that echo the mode of an operation as the operation is executed.

CHAPTER 3

Icons: The Graphical Counterpart of Earcons

Computer scientists, graphic artists, and human factors specialists have compiled a wealth of information concerning the structure, design, and use of icons. Valuable insight on the design of earcons can be obtained by investigating icons since they are the graphical counterpart of earcons. As with any new area, the most effective applications of earcons are still unknown. At the present time, the use of earcons can best be understood through the similarities of icons. This is not to say that earcons are identical to icons in both composition and usage. The analogy of earcons to icons is drawn primarily as a model for a new concept. This chapter takes a look into the world of icons; what they are, the principles behind their structure and design, and how they are learned and used by the computer user population.

3.1. Introduction

In the early period of computer interfaces, information was typically displayed in text. This trend has changed considerably within the last few years as more and more display information is represented by graphical symbols to convey information [Eas70]. These graphical symbols are known as icons and have become an important part of the display generated by a system. The success of iconic communication is that it uses imagery that relies on the human ability to quickly perceive natural form and shape [Hug74]. Contemporary computer systems utilizing icons in their interfaces are, for example, Xerox's 8010 Star Information System, Apple's Lisa and MacIntosh systems, Intran's Metaform, and PERQ'S Sapphire (Screen Allocation Package Providing Helpful Icons and Rectangular Environments).

A variety of computer entities can be represented in the user interface by icons. System state information, utilities, processes, programs, commands, cursors, menus, menu items, windows, screen selection buttons, and static objects such as a file are examples of various entities depicted by icons and their variety demonstrate the versatility of icons.

Star was the first system to use icons to represent windows [Smi82]. A temporarily unneeded window is shrunk down until only the icon representing it is visible. The icon for a window is a small picture representing its functionality. The icons in Lisa and MacIntosh not only represent windows but are also the actual selection buttons for the disappearance and reappearance of windows [Wil83, Mar84]. The MacIntosh also uses icons to describe available window choices. A picture of a scissors, for example, represents the cut and paste utility on an editing menu [Mar84]. The Metaform System uses pictures of pointing hands, brushes, paper clips, and pen and pencil points to represent the appropriate cursors to signal system processes and states to the user [Mar84]. Metaform also uses narrative images as icons to briefly explain the currant activities of a system module [Mar84]. Sapphire utilizes icons for window control (similar to that in Lisa and MacIntosh) as well as for multiple process monitoring [Mye84]. Each icon in Sapphire describes the process being ran in a particular window and the status of that window.

The principle use of icons in the Star, Lisa, and MacIntosh is to represent objects, operations, and operations on objects [Mye84]. Xerox's

Star first introduced icons primarily to make the operations on the computer seem more like the operations commonly found in the office environment [Mye84]. Its icons are, therefore, small pictures of the office objects and operations they represent such as in/out baskets, file folders, documents, and desk drawers. In the Lisa and MacIntosh, icons can also represent static objects , like a document, and moving one of these to an operation icon causes the process to operate on the object [Mye84]. Saving a file in a directory, for example, can be entirely performed via icon selection by putting a document icon in a folder icon in a file cabinet icon [Mar84].

3.2. Advantages

There are several advantages of representing certain types of display information generated by the computer with icons rather than written text.

- Icons can represent a lot of information in a small amount of space [Hem82]. "A picture is worth a thousand words." The most efficient and effective use of the display is important as space is a finite quantity on the computer screen.

- Icons make good distinctive targets on the computer screen as it is easier to spot a graphical image than a written word [Hem82]. Graphical symbols are also more visually distinct from one another than words [Hem82].

- Computer users can recognize and process graphical images faster than words. Icons have a definite visual resemblance to the object or operation it represents making mental processing less dependent on formal learning as evident in the processing of words which are very dependent on a specific language and culture [Hug74]. Furthermore, icons are easily learned, retained, and recalled as single units of information due to the powerful "parallel processing" ability of human image memory and processing capabilities [Gli84].

- Iconic menu selection is faster than words or phrases in menu selection [Mar84]. Icons allow the computer to communicate with the user by easily understood, nonlinguistic symbols which allow the user to respond to tactile movements that soon become second nature.

- Icons often have a universality about them that is not possessed by written language [Gli84, Kol69]. The use of icons in computer-user interfaces can further international communication between cultures speaking different languages. Computer systems are being sold and used by countries other than their producers. In order to facilitate and maximize this international exchange of computers, many systems require as much language free displays as possible. Apple took advantage of this in the development of the MacIntosh which is almost entirely free of the English language. It utilizes a universal symbol system to represent everything in or on the machine except for menus of single words that can easily be translated into other languages [Mar84].

3.3. Types

A study of icons in computer interfaces shows that there are different types of icons in use. Some are simple pictograms of familiar objects while others are stylized compositions of geometric shapes and marks. To simplify this investigation of icons, we will attempt to classify and distinguish between the different types. Borrowing terminology describing sign types in general [Mar84], we label these three types as (1)Pictorial (representational), (2) Symbolic (abstract), and (3) Picto-Symbolic (combination of pictorial and symbolic images).

Pictorial icons are typically simple pictures of familiar objects or operations. Many of the icons found in the Star, Lisa, and MacIntosh systems are pictorial icons (see fig. 1).

However, pictorial icons do have their limitations. Many objects and operations do not have a familiar or obvious pictorial representation. Even when they do, it is often difficult for a designer to draw and implement them since some objects and operations are detailed or complicated. Not every implementor is an artist! Symbolic icons are formed by combining geometric marks and shapes to depict a specific computer object or operation that is not easily or optimally depicted by a picture. The icons in figure 2 are examples of symbolic icons. Picto-Symbolic icons are composed of both pictographic and symbolic images. The icons in figure 3 are picto-symbolic icons.

3.4. Structure

The general structure of icons, regardless of type, can be broken into two components; elements and compounds [Kol69]. An element is the smallest interpretable unit and is a graphical image, either pictographic or geometric, whose decomposition would result in uninterpretable parts.

There are countless pictographic images and geometric shapes and marks that can be used as elements. A compound is any combination of these elements.

Pictorial icons can be composed of one or more pictographic images. Figure 1 are pictorial icons with (a) being composed of a single pictorial element and (b) being compounds of more than one pictorial element.

(a) Elements (b) Compounds

Fig. 1 : Pictorial Icons

Symbolic icons can also be composed of one or more geometric elements. Figure 2 shows symbolic icons with (a) being composed of a single geometric element and in (b) being compounds composed of more than one geometric element.

(a) Elements (b) Compounds

Fig. 2 : Symbolic Icons

The third type of icon, Picto-Symbolic, are compounds consisting of two or more pictographic and geometric elements.

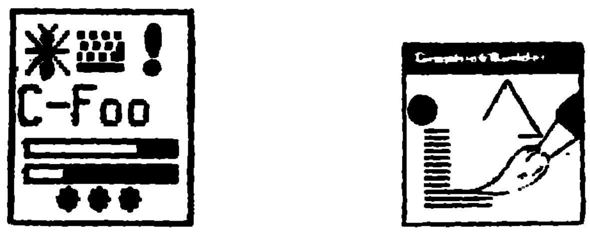


Fig. 3 : Picto-Symbolic Icons

Many computer entities representable by icons have a variety of features in common. Computer commands, for example, are highly structured sets with many commands sharing features with one another [Hem82]. The icons in figure 4 share the delete operation, represented by X in both icons.

Fig. 4

Composing icons sharing common features between entities can be done by using similar shaped elements to represent similar classes of information [Kol69].

The repetition of elements in icons that represent entities sharing common features has several advantages:

- It can facilitate initial comprehension and retention of the icon.
- The user has less to learn. When the user has comprehended one of the icons, he can apply what he already knows directly to a new icon.
- It is easier for the user to identify the features of the depicted object that corresponds to the features of the entity.
- New and unfamiliar icons are easier and faster to learn than those not sharing repeated elements. When icons for a familiar and unfamiliar command share an important element, the user is able to make predictions about the important characteristics of the unfamiliar command.

3.5. Design

The design of an icon is a specialized area of study that draws on many disciplines other than computer science, such as art and psychology. Pertinent information found in these various disciplines contribute to the formation of design principles for icons. Basic principles must exist as icons designed haphazardly using arbitrary or over-detailed images may be confusing to the user and inhibit man-machine communication.

The concept of "figural goodness" in Gestalt psychology helps in determining the general characteristics of a "good" icon. "Figural goodness" is described as:

The perceptual processes, in attempting to decode the incoming visual

> stimuli, examine the different structural patterns for clarity and stability. The more this perceptual process is enhanced by the inherent clarity and stability in the visual sensation supplied by the form, then the more "figural goodness" that form is said to possess. Thus if we wish to avoid confusion and ambiguity, we must have maximum figural goodness which in turn implies high internal organization and stability of the visual form [Eas70].

Based on this concept of figural goodness, Easterby [Eas70] describes five general characteristics possessed by a good icon. They are closure, continuity, symmetry, simplicity, and unity.

(1) Closure - Scientific evidence indicates the superiority of closed figures. The human perceptual mechanisms have a tendency to choose figures that help achieve high internal figural organization.

(2) Continuity - The continuity of an icon has a strong influence on its perceptibility. Icons with a smooth continuous outline are figurally good.

(3) Symmetry - Icons that are symmetric naturally facilitate its organization in the visual perceptual process.

(4) Simplicity - Icons should be as simple as possible. Fine details make no contribution to unambiguous and rapid interpretation of the icon by the user.

(5) Unity - Icons should be as unified as possible. This can be achieved by consistent use of the same size and positions of individual elements when they are repeated. When solid and line outline figures occur together, more unity if achieved when solid figures are integrated by enclosing them within line outline figures.

These guidelines of figural goodness are of particular interest to us as they apply to musical figures as well [Dav78].

Principles of good graphic design also provide information on the design of a good icon. Since icons are graphical images, designers may follow these principles in the creation of icons. Marcus [Mar84] describes graphic design principles in the concept of "corporate graphics". He describes this as:

> The concept of corporate graphics implies that all images are designed to meet unique communication needs, while being adjusted to produce a visual consistency with the system. This combined approach can be achieved by the use of a constant scale, limited size variations, the orientation of figures with respect to text, limited use of colors, limited variations in line weights, and the treatment of borders for figures or pictograms. These visual themes help to establish recognizability, clarity, and consistency...

Marcus believes that as the field of icon design grows and adopts the principles of corporate design and eventually standards, that all designers, implementors, and users will benefit "by being able to see and learn from successful approaches to typography, symbolism, color, layout, and sequencing" [Mar84].

The above principles provide valuable information on the general design of icons but they are somewhat vague for the actual design process. Easterby [Eas70] advocates a specific design approach to create icons, namely an experimental approach. This suggests that the designer first define a prototype for each icon in the interface based on the above general principles, what function of to be represented, and the notions found in the concepts of pattern perception and discrimination. The next step is to validate these prototypes through human factor experimentation on the intended user population. The final form of each icon is achieved by manipulating these prototypes according to the results obtained in the human factor experiments.

Each type of icon has unique design guidelines associated with it. The success of icons depends upon the experience of the user and his ability to infer the object or operation from the parts shown. Thus, the selection of pictorial elements plays an important role in the success of pictorial icons [Kol69]. The majority of pictorial icons are depictions of objects and/or operations that are familiar to most users. A good pictorial icon must use abstract elements to stress the generic qualities of the entity, yet it must be recognizable as a visual representation of a specific kind of computer entity [Sin83]. The best design of such icons is to represent the familiar object or operation by a canonical view of the item that shows or even exaggerates its distinctive features [Hem82].

There are a number of considerations in designing symbolic icons. The larger the icon, i.e. the more elements used in one icon, the more information it can display, but the harder they are to fully interpret and the less screen space is available. Marcus [Mar84] suggests carefully limiting the number of elements in a symbolic or picto-symbolic icon to a set that can be combined into the different forms required to depict the various functions of the system.

3.6. Interpretation

Recognizing even a realistic picture of an object requires a good deal of perceptual learning, abstracting ability, and intelligence. Icons too have to be read and their successful reading depends upon the user's knowledge and skill. To serve their purpose, icons require that the user acquire some special learning [Kol69].

The mechanism for reading icons requires that the user be able to abstract from his experience with the real object some relation of its distinguishing marks and to generalize the relation to the picture itself. Consequently, the icon need not be as detailed as the actual object because human memory is not photographic. Pictograms that emphasize distinguishing features make for easy recognition [Kol69]. One characteristic of pictorial representations in general is that the picture substitutes a part to indicate the whole [Kol69]. This makes it very easy for the user to decipher the intended meaning of that icon.

Human beings have evolved so their ability to process iconic messages is highly developed. This ability depends on the registering of visual images and upon the ability to retrieve specific icons up for review when needed without interference from other, perhaps quite similar, icons [Hug74]. An iconic message is apprehended by means of parallel processing. Icons are processed as complete units which allows the message to be rapidly and easily encoded by a user's perceptual processes. Evidence is mounting that the capacity for processing iconic information exists in amounts and for time periods that are quite large [Hug74].

PART II.

LEXICAL:

The Components of Sound

CHAPTER 4

Psychoacoustical Characteristics of Sound

Over the years, sound has been thoroughly studied by two different branches of science. Physics has examined sound according to the physical aspects such as frequency, intensity, wavelength, amplitude, phase, and duration. The science of psychoacoustics has studied sound in relation to how certain psychophysical features of auditory phenomena are perceived and experienced by their listeners. Psychoacoustics explores the subjective psychological dimensions of sound. These include pitch, timbre, rhythm, loudness, consonance, and dissonance. Each of these attributes contribute to the overall perceptual experience a listener receives from an audio cue. Twenty different attributes of sound have been identified [Yeu80], although only eight are discussed below.

Both the physical and psychological dimensions of sound are of importance and one should be aware of the distinctions between them. The earcon designer must be aware of the physical attributes so that the sounds composing the earcons can be created in a scientific, systematic, and predictable way. The human factors specialist must be aware of the psychological aspects of the overall effect the earcons have upon the ultimate user. Are they effecting the user in the manner intended? Are they detrimental to the user? It must be remembered that earcons are created to communicate information and they are not merely to entertain. Communication is achieved using both the physical and psychological properties of sound as well as other factors to be examined in later chapters.

In order to fully understand the guidelines given in this chapter, the physical dimensions of intensity and frequency must be defined. Intensity refers to the vibrational aspects of the sound wave and is expressed in terms of physical energy or pressure. The frequency of a sound is the number of cycles that a sound wave completes per unit of time. The frequency range of "optimal" human hearing extends from 20 to 20,000 Hz [Gel53,Sch76,von66]. However, by the age of thirty the average person hears no higher than 17-18,000 Hz; by fifty, up to 14-15,000 Hz; and by seventy, only up to 9-10,000 Hz [Ger85]. Our hearing is most sensitive to frequencies ranging from 200 to 4000 Hz [von66]. Studies have shown [Sch76] that the normal human ear is able to detect a change in frequency of about 3 Hz between two successive tones for frequencies up to about 2000 Hz when the tones are played at a moderate intensity level (at 40 db) [Sch76]. Broadbent [Bro54] has shown that human response to high frequencies is faster than to low frequencies. Generally, as a sound appears softer it is more difficult for the listener to detect it as different from other sounds close to it in frequency.

It is the purpose of this chapter to explore some of the psychoacoustical attributes and to specify how each effect the design and perception of earcons in a sound interface. Many of the most interesting phenomenon of hearing have to do with the interaction of sound that are heard simultaneously. Beats and masking are in this category and are discussed below. The information given in this chapter forms a basis upon which the guidelines in the following chapters are built.

4.1. Pitch

Pitch is a psychological sensation aroused by the physical stimulus of frequency to the human ear. As a result, low frequency tones are typically perceived as being lower in pitch than high frequency tones. Pitch represents how high or low a sound appears. Many people have been inclined to equate pitch with frequency because pitch is primarily determined by the frequency of a tone reaching the ear. This is due to the fact that high pitch sounds are typically heard from high frequency tones and low pitch sounds are heard from low frequency sounds. However, pitch and frequency are not synonymous and the two concepts should not be confused. Pitch is the psychological attribute that corresponds to the physical attribute of frequency. Most listeners can perceive the pitch of a sound after very few periods of the sound wave have been presented to the ear [Ras82]. The psychological dimension of pitch is scaled by the standard unit called a mel. By definition, the pitch of a 1000 Hz tone at 40 db is assigned a value of 1000 mels [Sch76].

M *An increase in the intensity of a high(low) frequency tone being used as an audio cue requires decrease(increase) in its frequency in order for it to remain at a constant pitch* [*Sch76*].

M *The pitch of an earcon should be within the frequency limits of 1000 to 4000 Hz for best pitch constancy* [*von66*].

4.2. Timbre

Timbre is a tonal attribute that refers to the quality or "richness" of a sound. Musical instruments sound different because the physical makeup of their sound differs. For example, the same tone played on a piano sounds different than when it is played on a clarinet. This is because each of these instruments has its own peculiar set of waveforms and overtones/undertones in addition to whatever fundamental pitch it may be sounding. The timbre of a sound is usually described with imprecise adjectives such as bright, warm, harsh, hollow, twangy, or brassy. Even though timbre is almost impossible to precisely describe and notate, it is one of the most immediate and easily recognizable characteristics of sound [Ker80].

The timbre of a sound is determined primarily by its "waveform". When a note is played, it consists of a sine wave oscillating at the fundamental frequency and the harmonics of that wave. Fundamental frequency is the basic rate of vibration of a sound wave and it defines the overall pitch of the note. Harmonics are sine waves having frequencies which are integer multiples of the fundamental frequency. A sound wave is the fundamental frequency and all of the harmonics it takes to make up that sound (see figure 5) [Com82].

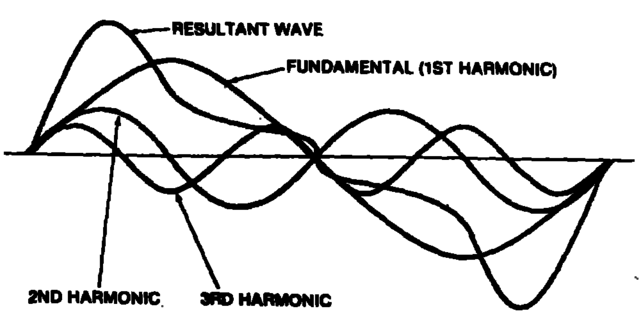


Fig. 5

In musical theory let's say the fundamental frequency is harmonic number one. The second harmonic has a frequency twice the fundamental frequency, the third harmonic is three times the fundamental frequency, and so on. The amounts of each harmonic present in a note gives it its timbre.

A triangular wave contains only odd harmonics. The amount of each harmonic present is proportional to the reciprocal of the square of the harmonic number. In other words harmonic number three is one ninth quieter than harmonic number one. Sawtooth waves contain all the harmonics. The amount of each harmonic present is proportional to the reciprocal of the harmonic number. For example, harmonic number two is half as loud as harmonic number one. The square wave contains odd harmonics in proportion to the reciprocal of the harmonic number. Other rectangular waves have varying harmonic content. By changing the pulse width, the timbre of the sound of a rectangular wave can be varied tremendously.

Musical examples of these relationships are the following. The biting, edgy sound of the oboe consists mostly of triangular waves. The smooth, sweet sound of a clarinet is mostly rectangular waves. Finally, the dry, clean sound of a flute consists almost purely of sine waves.

A sound wave also vibrates in a series of parts where each part represents an overtone, or partial. The timbre of a sound is primarily a function of the number of overtones or partials that the sound happens to have as part of its vibrating frequencies. Most sources of sound, including musical instruments and the human voice, produce sounds composed of a series of overtones.

The intensity of the fundamental and partials also affects the quality of sound. In general, the greater the intensity, the more partials will be present. Also, lowering or raising the frequency of the sound alters the timbre.

P *The sound and perception of an audio cue can be altered by changing the waveform or timbre of that audio cue [Gre85].*

4.3. Rhythm

Rhythm, in the most general sense, is the term referring to the whole time aspect of music. In a more specific sense, a rhythm or the rhythm of a certain sequence of tones refers to the specific arrangement of long and short notes in the sequence and their accents [Ker80].

Listeners often respond to music by participating physically in its rhythm through tapping of feet because they are carried along by the characteristic of rhythm. Rhythmic pulse or "beat" can be compared to the tick of a watch or heartbeat. It is the shortest, most easily felt time division in a given segment of music. The human ear responds easily to pulse or beat when it is periodic or regular.

Note Type	Notation	Duration *
1/16		128
1/8		256
Dotted 1/8		384
1/4		512
Dotted 1/4		768
1/2		1024
Dotted 1/2		1536
Whole		2048

* Time in milliseconds

Fig. 6

In music, rhythm is indicated with note values. Common time divisions and their iconic notations are listed in figure 6. A whole note is given the longest time value. A half note lasts for half the time of a whole note, a quarter note lasts for a quarter of the time of a whole note, and so on.

P *Short silences as well as timed sounds can be used to create rhythms for audio cues.*

4.4. Loudness

Loudness is a psychological attribute of audition that corresponds closely to the physical attribute of intensity. Loudness is primarily determined by the physical intensity of a sound but it is also a function of the frequency of a sound. Therefore specifying only the decibel level of a sound will not fully describe its loudness. Loudness is not identical to intensity and decibels are not a measure of the experienced loudness. Loudness is measured by a unit called a sone. A sound is defined to have the loudness of one sone if it is as loud as a pure tone of 1000 Hz at an intensity of 40 db [von66]. Loudness is also a function of the duration of the tone. In general, tones lasting less than one half of a second appear less loud than tones of the same intensity but of longer duration [Lun53].

M *In order to double the loudness of a audio cue, the intensity must be increased by 10 db [Sch76].*

M *In order for a low frequency tone to sound as loud as a middle frequency tone in an earcon, the intensity of the low tone must be increased [von66].*

An 80 db 25 Hz tone sounds as loud as a 40 db 1000 Hz tone.

M *The loudness of an audio cue can be calculated by the following formula:*

$$L = k * I^{0.3}$$

where

L = *psychological dimension of Loudness*

K = *a constant*

I = *physical Intensity*

This says that the loudness of a sound increased approximately as the cube root of the sound intensity. This equation indicates that loudness grows more slowly than physical intensity. It also shows that increases in intensity produce lower proportional increases in loudness [Sch76]. Other exponents have been reported for the scaling of loudness. They are:

0.54 for monaural loudness [Sch76]

0.67 for binaural loudness [Sch76]

0.60 for the musical scales [Mat69]

M *An audio cue whose frequency and intensity lie on one of the curves in figure 7 appears equally as loud as any other audio cue on that same curve although their frequencies and intensities differ [Pie58, Sch76].*

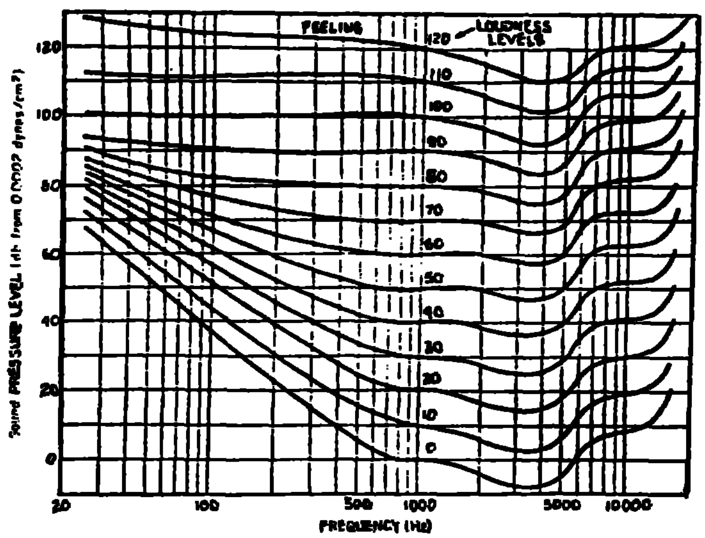


Fig. 7

4.5. Consonance and Dissonance

When two tones are played simultaneously, the resulting combination often consists of sound relationships that are either simple or complex.

Those with simple sound relationships are called consonant and those having more complex sound relationships are called dissonant. The terms consonance and dissonance are often used to describe sounds appearing pleasant or unpleasant to the normal human ear. However, since pleasant/unpleasant imply a qualitative judgement that differs from person to person, this discussion will abandon their use for a more scientific approach to this relative issue.

Many auditory characteristics as well as non-auditory characteristics such as culture, custom, and musical preference play a role in determining if a sound is consonant or dissonant. One of the major auditory characteristics that determine consonance or dissonance is the frequency difference between the tones involved. People typically find pairs of tones whose frequencies are related 2:1(the octave), 3:2(the fifth), 4:3(the fourth), 5:4(the major third), and 6:5(the minor third) as consonant. They find pairs of tones with frequency ratios that fall between these intervals to be dissonant [Mat69].

M *Earcons will appear to be dissonant when two strong partials fall within the critical bandwidth [Mat69].*

The critical bandwidth of the frequency of a sound is defined to be about 100 Hz below 600 Hz to about a fifth of an octave above 600 Hz [Mat69].

M *Earcons composed of complex tones are consonant when the low tones are separated by a larger fraction of an octave than the high tones [Mat69].*

4.6. Beats

When two tones that have slightly different frequencies but are of similar intensity are played simultaneously, the waves of these two tones will sometimes reinforce and at other times interfere with each other. When they reinforce each other they produce an increase in the heard intensity. When they interfere with each other, they produce a decrease in the heard intensity. These fluctuations in intensity and perceived loudness are called beats.

The auditory experience produced in beating is greatly affected by the rate of the beats. The beating rate becomes faster as the frequency difference between the two tones involved increases. Experiments have shown [von66] that minor slow beats do not bother the listener very much but faster beats are inclined to cause the listener to become anxious or irritated. This can be compared to the annoying flickering that a movie projector produces when its speed is slowed down too much. A listener is no longer able to discriminate a fast beat (one whose tonal frequency difference is greater than 20 Hz) as an individual entity any longer. Instead it will sound like a continuous, harsh dissonant sound. Even though beats are very hard to hear without training and as such they would not be a very useful element in differentiating earcons from one another, they still can be used to add interesting audio dimensions to audio cues.

M *When beats are included in an earcon, the frequency with which the loudness will fluctuate is precisely the difference between the frequencies of the two tones combined* [*Sch76*].

M *The frequency difference between the pairs of tones composing a beat in an earcon should be within the range of 0-20 Hz* [*Ras82*].

4.7. Masking

Masking is a phenomenon that occurs when two tones are played simultaneously and one of the tones drowns out the sound of the other tone. It is said that the more powerful tone masks or disables a listener from hearing the less powerful tone. Masking occurs when the two tones involved are close in frequency but not in intensity. The masking tone has greater intensity thereby reducing a listener's perception of the softer tone. Tones close in frequency to the masking tone are more strongly masked (drowned out) than those far removed in frequency. The effect of masking is greater on tones whose frequencies exceed the masking tone than those with frequencies below it. Low tones have a greater masking effect than high tones. Also a greater amount of masking results when the masking tone is more intense.

When designing earcons, it is necessary to know if and when the tones chosen for a particular earcon mask each other or not. Depending on the task situation, masking may or may not be desirable. What is important is to know how masking occurs and under what circumstances will it be present.

P *Tones that are required for information in an audio cue should not be masked.*

PART III.

SYNTACTICS:

Formal Arrangements of Sounds

CHAPTER 5

Arranging Sounds into Earcons

The design of an icon and earcon have similar communication needs and design problems in spite of their obvious differences in that they utilize different senses. Successful approaches to icon design are now well established and have stood the test of time. The knowledge obtained through an examination of the theories behind design schemes for icons can be used to model equally successful approaches for earcons. In this chapter, a systematic design approach for earcons will be presented.

Just as in icon design, many different schemes for designing earcons are possible. One such scheme might be the use of digitized natural sounds from our surrounding environment as earcons. The first inclination with vision is to use highly representational pictorial images as icons. In the same way, highly representational natural sounds could be used as audio cues. The cue for *filling* in a graphical area, for example, could be the digitized sound of a water faucet. It should be noted that it is not the intention of this thesis to imply that the design approach suggested is the only or even the best possible approach; it is one plausible method.

The following approach to earcon design is modeled after Marcus's [Mar84] approach to icon design as well as after the theories of music composition. Marcus describes a set of "elements", geometric shapes and marks, to be used as building blocks to construct a set of icons. The power behind his method is in the very existence of these well defined elements which can be put together in various ways to build large and highly customized sets of icons. In the same way, the earcon design approach advocated here uses single pitches and motives as its "elements" or building blocks for earcons. Motives, described more fully in the following section, are sequences of pitches that create a short and distinctive audio pattern often characterizes by the simplicity of its rhythm and pitch design [Chr66]. Their very brevity, coupled with its distinctive manner, enable motives to play a powerful role in the composition of earcons.

The motive was chosen as the main building block for earcons because it fulfills all of the same requirements of its graphical counterpart, Marcus's geometric shapes and marks. Like Marcus's elements, motives are well defined. They have been used for centuries as the smallest unit in music composition [Ber66]. Also, motives can easily be used to build large and customized sets of earcons.[1]

This particular approach to earcon design was chosen for several reasons:

(1) It is a systematic approach having well defined building blocks. A systematic approach is usually straightforward to understand and use.

(2) It retains all of the power and advantages behind Marcus's approach for icons, specifically, how elements are used to easily build larger sets of objects.

(3) The existence of well defined building blocks brings to this approach all of the aspects and benefits of modularity. Modularity allows easy

[1]The reader may question that complex earcons can be easily assimilated due to the transient nature of sound. However, using principles well understood in musical composition, we will show that earcons as well as icons may be systematically constructed out of basic units.

modification, future set expansion, and tailorability of a set of earcons.

(4) It is easily implementable on most contemporary computer systems, large or small. Most computers today have the necessary equipment to perform rhythm and simple pitch. This approach makes it unnecessary to invest in expensive sound boards or other specialized equipment.

Regardless of which approach is taken for the design and implementation of audio cues into a sight and sound user interface, ***the incorporation of an expert on "sound" into the design team of a sight/sound computer-user interface is of utmost importance***. The science of sound is a highly technical, diverse, and complicated discipline. Only an expert in this field understands the existence, importance, implication, consequence, and avoidance of the many perceptual problems and intricacies of sound.[8]

5.1. Motives

A motive is a brief succession of two or more pitches arranged in such a way as to produce a tonal pattern sufficiently distinct to allow it to function as an individual recognizable entity [Ber66]. Examples of motives in musical composition are shown in figure 6.

[8] Robert Greenberg, a Ph.D. in music composition (University of California at Berkeley, 1984) served as technical advisor for this thesis. Most of the technical information and suggested guidelines presented in this chapter were developed during numerous consultations between Dr. Greenberg and the author.

Bach: Brandenburg Concerto No. 3, I.

Fig. 8

Rhythm and pitch are the two elements of motives. Three characteristics that become important for differentiating a given motive in earcon construction are timbre, register, and dynamics. As demonstrated throughout the remainder of this chapter, a single motive can be easily modified to create very different and distinct sounds by varying any one or all of these five characteristics.

The eloquence of motives lies in their ability to be combined to create larger recognizable structures [Chr66,Ber66]. Repetition of motives, either exact or varied, or the linking of several different motives produce larger, more self-sufficient patterns [Chr66,Ber66]. These larger structures are earcons. As motives are put together to form these audio cues, they can have the following useful relationships with each other [Rat66]:

(1) Repetition - exact restatement of a proceeding motive.

(2) Variation - some exchange but not enough to disguise the similarity to a proceeding motive.

(3) Contrast - decided difference from the proceeding motive.

These relationships play a key role in the composition of earcons.

5.2. Motive Construction

The design of a motive involves choosing a specific value for each of its five characteristics or parameters; rhythm, pitch, timbre, register, and dynamics. It is vital to differentiate between the primary aspects of pitch and rhythm - the "steel and concrete" of motive construction - and the secondary aspects of timbre, register, and dynamics - "the paint and wallpaper" - which help listeners to distinguish between otherwise like motives. Timbre, register, and dynamics alone or together cannot create a motive. They can only decorate one. By definition a motive is a "rhythmicized sequence of pitches" - rhythm and pitch.

Although a large number of values are possible for each of these parameters (for example, there are countless values for rhythm as time can be infinitely divided), limiting the available selection choices to a small and workable subset has certain advantages. It simplifies motive construction by restricting parameters while still allowing wide freedom in design style and choices of sounds for earcons.

Parameter	Total Choices	Description
Rhythm	4	Dividing 1/4 notes into 1/16 notes.
Pitch	96	8 octaves of 12 pitches each
Timbre	4	Sine, Sawtooth, Triangular, Rectangular
Register	3	Low, Medium, High
Dynamics	5	Soft, Medium, Loud, Soft to Loud, Loud to Soft

Table 1

The numbers of choices for each parameter suggested in Table 1 is one scheme for limiting selection values. Using the "Basic Principle of Counting" in Combinatorial Analysis [Ros76], this set of limitations still allows for the creation of over five billion unique two to four note motives! This is obviously more than enough sonic information from which to construct useful audio cues.

The sections to follow provide information and guidelines concerning the specification of each of the five characteristics in the construction of a motive.

5.2.1. Length of Motives

M *The optimal number of pitches in a motive used in an earcon is two to four pitches.*

Four pitches as an upper limit is suggested for several reasons. Sound takes place in time so, unlike vision, a listener can't dwell on a sequence of pitches. It is therefore best to keep motives as short as possible so listeners can make the necessary audio connections between the pitches they hear. The use of very short motives in earcons alleviates many of the inherent perceptual and musical problems of sound. Motives consisting of greater than four pitches can have undesirable melodic implications. A long succession of pitches tend to hierarchize themselves creating a melodic pattern.

To better understand the following material, it is helpful to define and contrast the words melody and tune . A *melody* is any succession of pitches heard in association with one another. Therefore, motives are melodic constructs. By contrast, a *tune* is a melody that becomes, by simplicity of construction, phrase structure, and contour, instantly recognizable and singable.

As soon as a motive in an earcon sounds like a tune, the listener will associate that earcon with music. Hearing a simple tune ten or more times a day could potentially be irritating to a computer user as well as lead to audio fatigue. Motives consisting of four or fewer pitches are long enough to represent computer information adequately via an earcon but short enough to avoid a specific tune and user fatigue.

5.2.2. Rhythmic Structure

Rhythm is the most prominent characteristic of motives. A listener more easily remembers the beat or rhythm of a song over other characteristics such as lyrics or tonal composition. Choosing a rhythmic structure means specifying the timing of each pitch in the motive. There are many possible different rhythmic structures available to create different and distinct motives when using only the time values described in figure 6 (section 4.3).

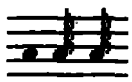
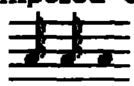
Let the total amount of time required for the execution of each note in an earcon be referred to as the time complexity of that earcon.

P *An earcon should have time complexity as low as the comprehension limitations allow [Bri78].*

In other words, the earcon should have a total time length long enough to get the message across effectively, but no longer. Have the time unit fit the message, not visa versa. The earcon's meaning must come across as quickly and as easily as possible since sound is transient.

M *A user's attention decreases as the length of the earcon increases, all other factors being equal [Bri78].*

P *The easiest way to produce two distinct motives is to vary the rhythmic structure of each motive.*

The same sequence of pitches can be made to sound differently when different rhythmic structures are imposed on it. For example, changing the rhythm of  to  produces dissimilar sounds using the same three pitch sequence.

5.2.3. Pitch Selection

P *The pitches in a motive should be specifically and carefully chosen.*

The pitch sequence of motives can take a variety of forms due to the large number of available pitches from which to choose them. Most computer users are familiar with the tonal system of eight octaves of twelve pitches each. When only considering these 96 pitches, there are 96 possible unique one pitch earcons; 9,216 possible two pitch motives; 884,736 possible three pitch motives; and 84,934,656 possible four pitch motives. Although a large variety exists, motives constructed of random pitches from these 96 available is undesirable. Which pitches are chosen for a particular motive is important. For musical and perceptual reasons, certain pitch sequences don't sound satisfying or pleasing to the listener. The precise reasons are complicated and best understood by a person knowledgeable in music theory.[3]

P *All pitches in a motive should be chosen from the same octave.*

To avoid problems inherent in octave perception, it is best to compose motives for an earcon using pitches all belonging to one octave. This also facilitates changing the register of that pitch sequence, as seen in Section 5.2.5.

P *The simpler the entity or task being identified by an earcon, the simpler the harmonic makeup of the motive should be, i.e. all pitches in such an earcon/motive should have related overtones.*

The reason a chord sounds satisfying is because it is composed of notes having common overtones. Therefore, a motive composed of pitches having related overtones will also sound satisfying to the listening computer user. Staying close to the harmonic relationships of pitches avoids undesirable musical implications thereby alleviating many problems.

In order to understand the following important information and guidelines on motive construction, a short digression into musical theory is necessary.

[3] This is a prime example of the importance of including a music expert into design team for a sight and sound computer-user interface.

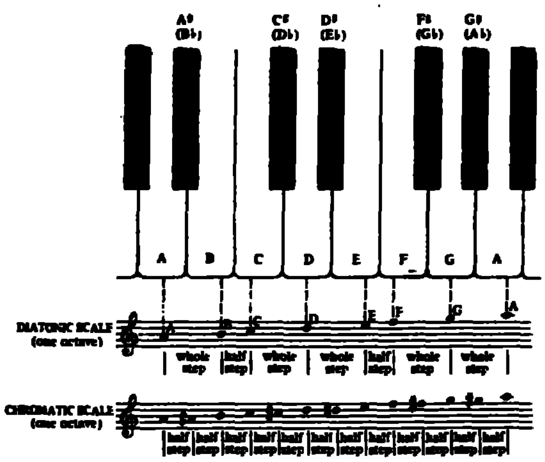


Fig. 9 : An Octave

The octave is divided into twelve equal segments or twelve pitches. These twelve pitches are collectively called the chromatic scale. The distance between any adjacent pitch in the chromatic scale is called a half step. For example, C - C# and B - C are each a half step apart. Two half steps equal one whole step. C - D and F - G are each a whole step apart. The existence of half steps in a sequence of pitches creates musical tension which should generally be avoided in the context of motives for earcon construction [Gre85].

P *For maximum clarity, earcon motives should be based on the preexisting pitch structures known as the diatonic major/minor scales.*

P *Great care must be exercised in the construction of earcon motives, particularly in terms of the use of half steps in those earcon motives.*

Half steps should be treated very carefully as their misapplication can create musical tension - a sense of non-resolution - and therefore musical implications which will contradict and interfere with the purpose of the earcon. For example, suppose a C is tonic (the one particular pitch that sounds fundamental or central to a piece of music) in a motive and the motive completes on a B. If the entire motive composition is such that there us a sense of priority about this C, a sense of non-resolution will ensue, as if it should have ended on a C but didn't. The musical reason that this motive sounds incomplete is because the ending B is only a half step from the tonic C. Here, the mere presence of the B - C half step at the end of the motive created an urgency. Once a pitch priority is established (C in this example), there might be a sense of urgency to go back to the original pitch.

P *If a pitch priority is established in a motive/earcon, it should be satisfactorily* resolved over the course of that motive/earcon.

M *Musically neutral sequences of pitches are best in motive construction for earcons.*

Musically neutral pitch sequences are achieved by avoiding as much as possible the intervals in the major scale that create strong tendencies, i.e. those involving half steps.

This is not to say that half steps can't ever be used, only that caution must be exercised if they are.

5.2.4. Timbre

The tonal quality, or timbre, of a motive is determined by its "waveform".

P *It is best to limit the timbre of a motive to those created by one of the four basic waveforms; sinusoidal, sawtooth, triangular, or rectangular.*

The sinusoidal waveform (pure unmodulated sound) and the sawtooth waveform (a jagged sound) are the most obvious waveforms but the triangular and rectangular waveforms are two other equally useful waveforms for motives. Although other possibilities exist as these four waveforms can be combined in various ways, it is best to stick to these four waveforms and the resultant timbres.

P *A very easy way to alter our perception of a given motive is to simply change that motive's waveform thereby producing a different timbre.*

Changing the waveform of a motive from sawtooth to triangular, for example, produces a sound quality that is different, less twangy, more hollow [Com82].

For a more information on timbre and how it is produced, see Chapter 4.

5.2.5. Register

Register refers to the relative high/low of a set of pitches and it describes their octave location. A low register motive sounds "low" and a motive with high register sounds "high". Labeling the eight octaves in our tonal system one to eight, from low to high respectively, a motive with low register is one in which its set of pitches are entirely contained in one of the lower octaves - octaves one or two. Likewise, a motive with medium register has pitches selected from octaves four or five, and one with high register has pitches selected from octaves seven or eight. Register can be creatively manipulated to indicate vertical location or direction in a sight/sound interface.

p *Motives composed with pitch sequences contained in the low, medium, or high octaves are easiest to remember.*

P *A motive can be made to sound very differently by changing its register (low, medium, or high) to a different value.*

Section 5.2.3 on "Pitch Selection" suggested that the set of pitches composing a motive be chosen from the same octave as this makes it very easy to change the register of that motive, thereby producing a new sound.

5.2.6. Dynamics

Dynamics is the relative loud/soft of a motive. Two obvious ways to differentiate within relative loud/soft is to have the dynamics of a motive be either constant or variable. This produces the four basic variations:

(1) loud for the duration of the motive

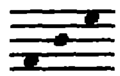
(2) soft for the duration of the motive

(3) a gradation from loud to soft

(4) a gradation from soft to loud

However, any combination of these four is also possible.

The parameter of dynamics can be creatively used in motive to indicate direction from left to right. For example, the three note motive



Soft –> Loud

whose dynamics changes from soft to loud could be used to indicate a left-right direction or movement.

The musical term describing this feature of dynamics is the crescendo. Crescendo means "to grow" in Italian and is a sound going from soft to loud. The opposite, a decrescendo, means "to get smaller" and goes from loud to soft. Both crescendo and decrescendo can be used in motives to indicate location, for example, of a window on the computer screen. Also, an earcon representing the operations of enlarging, zooming, and shrinking could be constructed from motives creatively varying the parameter of dynamics.

5.3. Earcon Construction

Following Marcus's [Mar84] scheme of building icons from a set of graphical elements, earcons can also be built by combining its audio "elements", a single pitch and motives, in various ways to create a set of earcons to represent various user interface entities.

Just as icons were of different structures which coexisted in the same user interface, so are earcons. Combining one or more audio elements to create an earcon leads to two basic types of structures; (1) earcons composed of only one element and (2) earcons composed of more than one element.

P *Choose the structure that most closely resembles the nature of the application and entity that the earcon is to represent.*

The point is that there is no one supreme structure. Each application has its own appropriate structure due to the close correlation between the nature of the entity and structure (see Chapter 7).

The following sections investigate each type of structure by discussing its characteristics, usage, and construction. An example earcon for each type of structure is given.

TYPE 1: Single Element Structures

Earcons of this structure are composed of only one element, either a single pitch or a single motive. Single element earcons are the simplest structure.

5.3.1. Single Pitch Earcons

Characteristics:

Any audio cue composed of one note with the attributes of pitch, duration, register, and dynamics.

Usage:

Since this structure is simplest, it can be used to represent simple, basic, or commonly occurring user interface entities such as key clicks, cursors, selection mechanisms.

Construction:

To build an earcon of this type, the following must be specified:

(1) A pitch to attach to the singular note. The octave location of this note automatically specifies the register.

(2) The duration of time that the note is to be played.

(3) The dynamics for the duration of the note.

Example:

A single pitch whose dynamics change through time from loud to soft can be used as an audio cue to indicate the system going down. A single pitch suffices nicely here because a falling dynamic sounds like it is actually going down and is therefore a realistic audio-pictorial representation of the application it is depicting. Another example of a one pitch earcon is a one pitch key click representing an editing mode, say the delete mode. The user hears the pitch each time a character is deleted.

5.3.2. Single Motive Earcons

Characteristics:

Any earcon composed of only one motive having the attributes of rhythm, pitch, timbre, register, and dynamics.

Usage:

Single motive earcons are also of the simplest structure. They therefore can be used to represent basic and common computer entities such as certain error messages, system information, windows, and files.

Construction:

Since earcons of this type are constructed of one motive, an explanation on constructing earcons of this structure are the same as those presented in section 5.2 on "Motive Construction".

Example:

Fig. 10

The three note motive in figure 10 could be the earcon representing some system message in one of several windows currently on the display screen. The register and dynamics of this earcon can be manipulated to indicate the location of the window displaying the system message. Low register and soft dynamics, for example, could be used to indicate that the lower left window contains an system message to be noticed by the user.

TYPE 2: Multi-Element Structures

5.3.3. Compound Earcons

Characteristics:

Any earcon constructed by combining two or more audio elements in succession.

Usage:

This structure is used for composing earcons that represent those computer entities sharing common features.

Construction:

Composing earcons sharing common features between entities can be done by using similar audio elements to represent similar classes of information.

We must assume the previous construction of a set of audio elements, one element for each of the entities in the system that represent the common features, that can be combined into the different earcons required to depict the various functions of the system. The elements of this set, called the ***primary set***, are constructed according to the guidelines on single element structures for earcons.

The construction of a new earcon involves putting together audio elements from the primary set in various combinations (where each element is included for a specific reason) to represent a new system function or computer entity.

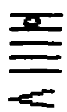
This method can be used to systematically build up a large set of earcons for the entire system. The repetition of audio elements to in earcons that represent entities sharing common features has the same advantages as it had for constructing icons, namely ease of construction, set expansion, and ease of user identification and retention (see Section 3.2).

Example:

Let the set of building blocks representing the common features consist of audio elements representing the computer entities CREATE, DESTROY, FILE, and TEXT STRING.

More specifically, each of these entities has the following compositions:

- CREATE represented by one pitch with dynamics going from soft to loud, i.e. a cresoendo, represented by



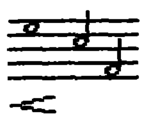
- DESTROY represented by one pitch with dynamics going from loud to soft, i.e. a decrescendo, represented by

- FILE represented by a two note motive with rhythm (long, long).

- TEXT STRING represented by a two pitch motive with rhythm (short, short).

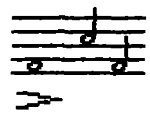
Using these audio elements for the entities above, we can systematically construct the earcons for the computer entities CREATE FILE, DESTROY FILE, CREATE TEXT STRING, and DESTROY TEXT STRING by combining the audio elements for the common features in the following way. The earcons for these larger functions could then be:

- CREATE FILE

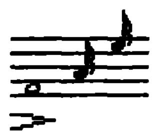


- CREATE STRING

- DESTROY FILE



- DESTROY STRING



5.3.4. Family Earcons

Characteristics:

Any earcon composed of a sequence of two or more motives each having characteristics in common such as rhythmic structure and length in terms of number of notes per motive.

Usage:

This structure imposes a useful hierarchy on earcons and the computer entities they represent. It can be used to construct earcons to depict families of computer entities by representing them with audio hierarchical earcons.

Construction:

(1) The hierarchy of computer entities must be described, i.e. the entity families and their descendents to be represented by audio cues of this structure must be identified. Examples of such families are the ERROR MESSAGE FAMILY, the WINDOW FAMILY, the PROMPT FAMILY, the WINDOW FAMILY, and the EDITOR FAMILY.

(2) Each family is assigned a different and distinct motive having a unique and identifying rhythmic structure. This motive, called the *family motive*, becomes the earcon that represents the family type, the audio version of the family's last name. Family earcons have no pitch structure; they are purely rhythmic sequences. Clicks or other noise sounds can be used to articulate their characteristic rhythms. Each entity at the highest level in a family hierarchy is represented by an earcon composed solely of the rhythmic family motive.

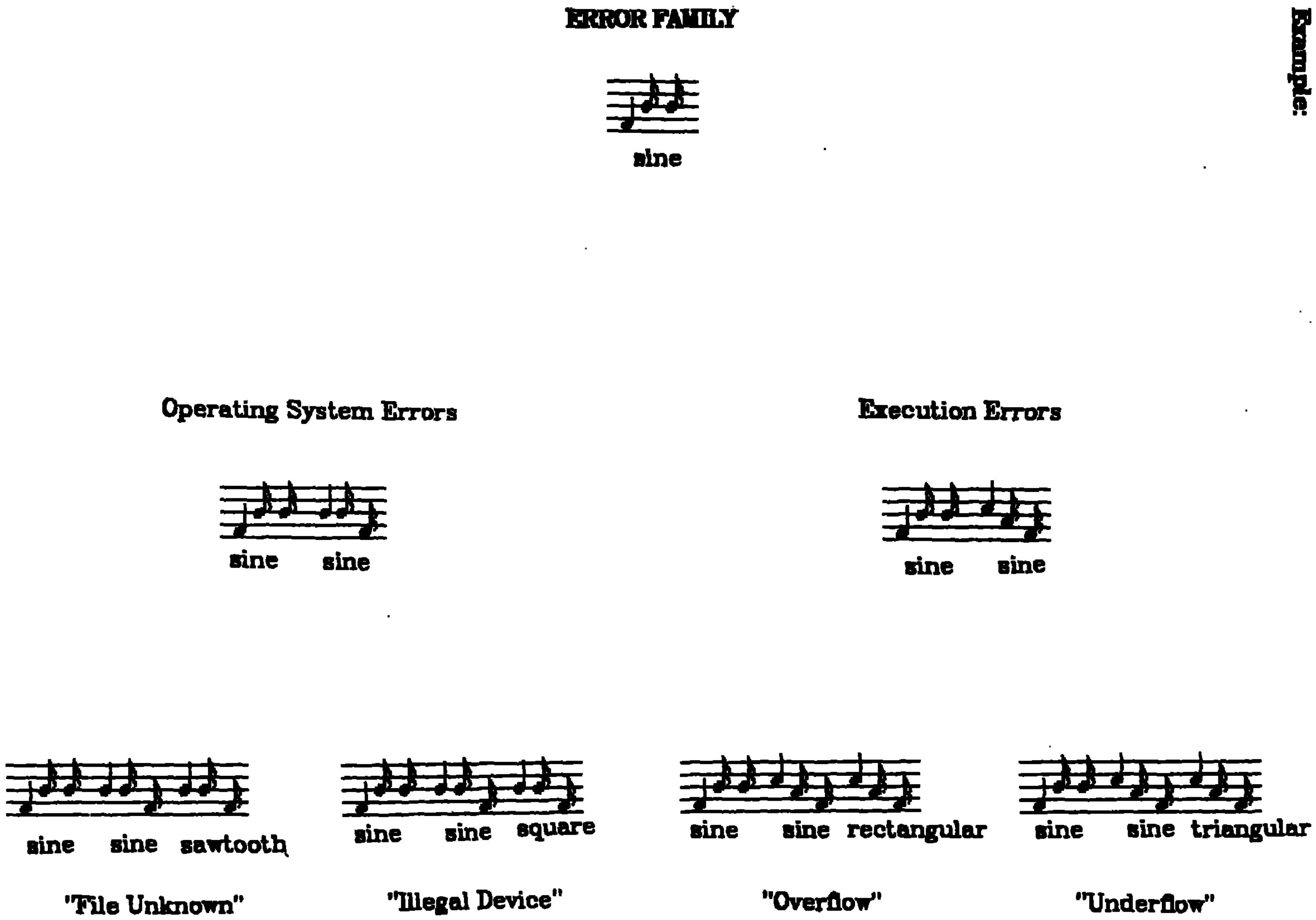
(3) All entities residing at the second highest level are represented by a two motive earcon. The first motive in this sequence is the family motive. The second motive is a duplication of the rhythm of the family motive, but is assigned a pitch sequence. This second motive is called level 2 motive. It is suggested that all level 2 motives use a sine wave as the timbral characteristic. Sine waves are the most "colorless" of the wave forms and this will help to avoid the difficulty of differentiating between the timbre of level 2 and level 3 earcons.

(4) All entities residing at the third level of the hierarchy are represented by a three motive earcon. The first and second motives in this sequence are those belonging to the two higher levels, namely the family motive followed by level 2 motive. The third motive is a duplication of level 2

motive but is assigned a different timbre. This third motive is called level 3 motive.

(5) If more than three levels exist in the hierarchy, the parameters of register and dynamics can be used to create earcons to represent them. For example, all level four entities could be depicted by the three motive earcon from level three but having different register or dynamics.

Example:


ERROR FAMILY
sine
Operating System Errors
sine sine
Execution Errors
sine sine
sine sine sawtooth
"File Unknown"
sine sine square
"Illegal Device"
sine sine rectangular
"Overflow"
sine sine triangular
"Underflow"

In each of the level three earcons, the first motive identifies the entity family (error, system information, menu, etc.), the second motive identifies the type of error (operating system, execution, syntax, compilation, etc.), and the third motive identifies the exact error message. This audio cue could accompany the textual error message appearing on the screen.

PART IV.

SEMANTICS:

The Meaning & Interpretation of Earcons

CHAPTER 6

Principles for User Comprehension

In order for any audio cue to be an effective component of a sight/sound interface, it must be well designed. It should convey one basic meaning, be brief, simple, distinct from other earcons, and be easy to remember, identify, and understand. It should also concern the listening user, deal with the current computer situation the user is in, and have some purpose for being there. Users will adjust and accept a new sound interface more quickly if they see of what use the audio cues are to them. It has been found that the more valuable an interface is in satisfying users needs and wants, the more likely that they will pay attention to and use the system [Bri78].

This chapter includes guidelines aimed at satisfying these goals. Any audio message, no matter what type it is, should adhere to these general principles. Otherwise all benefits from them will be negated.

6.1. Understandability

Understandability refers to an audio cue's purpose, meaning, and use being clearly understood by the user.

P *The purpose of an earcon must be absolutely clear and unambiguous.*

The user should know why a particular earcon is being played, what the system wants them to do, what they can expect in return, and what the consequences are of the present situation or action.

P *The meaning of the earcon must be absolutely clear and unambiguous.*

P *An earcon should have only one associated meaning.*

An earcon should not have multiple interpretations. Ambiguity lends to misperceiving by the user and perhaps failure of the user to understand the messages intended meaning.

P *The earcon should use sounds that are understood in the same way by the user as the designer intended them to mean [Bri78].*

See Chapter 10 for further information.

6.2. Distinctiveness

Distinctiveness refers to the psychological distance between an earcon and surrounding audio cues.

P *The distance between an earcon and surrounding audio cues should be maximized.*

Within the limitations of the parameters, the distance between two or more earcons should be maximized for easy user recognition. When composing earcons, choose distinct sounds and distinct motives which can be easily discriminated from surrounding noise or sound messages.

M *Minimal differences between different earcons, i.e. a small distance, increases confusion and therefore the error rate* [*Eng75*].

M *Confusability varies inversely with the meaningfulness of the distances between earcons* [*Hem81*].

Assuming we have a finite collection of n distinct sound elements from which the earcon is composed and assuming that the user does not perceive a melody or common pattern between the two earcons, we have the following guideline.

M *The distance between two earcons of n elements is defined as the number of elements in which they differ* [*Bla78*].

A large distance corresponds to the earcons being dissimilar and a small distance corresponds to the earcons being very similar.

The above is a measurable guideline used in coding theory. However, if the user does perceive a higher order organization such as a melody between the two earcons, it may no longer be a valid measurable guideline. In this case, the two earcons are considered to be audio translations of the same motive and are therefore very similar resulting in a small distance between them.

Keep different earcons distinct by choosing distinct sounds as well as distinct motives.

6.3. Consistency

Consistency refers to the property that similar audio messages are in agreement and conform to the same principles. The following two rules are stated to clarify the use of consistency in the context of this study.

(1) The same operation on an object always yields the same result.

(2) Similar operations on similar objects evoke similar results.

There should be consistency within an earcon and also between similar earcons.

P *Similar computer events should be represented to the user in a standard and consistent manner.*

For example, all error messages should be represented by the same type of audio cue.

P *Similar earcons should have similar meanings and be composed with similar sounds.*

This reduces user confusion.

P *The greater the consistency among the sound elements within an audio cue, the greater chances that the entire message will be perceived by the user as the designer intended [Bri78].*

P *An audio cue that is highly consistent with its associated computer event is more easily recognized and understood [Bri78].*

P *An audio cue's image and the tones composing them should be consistent.*

For example, urgent computer events should be represented by urgent sounding tones.

P *An earcon's sound elements and the user's affective response should be compatible or consistent.*

For example, if it is an "urgent" audio cue, then it should be composed of tones that produce urgent affective reactions in the listening user. See Chapter 10 for more details.

CHAPTER 7

Placement in Interface

7.1. General Information

An audio message is partly defined by the characteristics of the physical environment in which it is delivered. A user's perception of a message may be sharpened, distorted, obscured, or completely impaired as a result of its setting within the computer/user interface. Everything that a user hears and sees contributes to a total effect on the user, both consciously and subconsciously. The timing and placement of earcons is of utmost importance so that they result on a positive total effect that is of benefit to the user. Therefore, sight/sound interface designers must be aware of what they want the user to hear and see as well as how they want the user to react to what they hear and see. Everything in the sight/sound interface must have a meaning and reason for being there.

P *There must be precise timing between the visual computer screen and its supporting audio cues.*

Timing is of the essence or else all benefits may be negated, leading the user into confusion.

P *Audio cues must be strategically placed in the sight/sound interface as to best reflect the intended purpose of each earcon's presence.*

P *Audio cues should be placed as to minimize confliction with other current computer activities.*

Users are more likely to be exposed to an audio message if conflicting activities are minimized than if the user is distracted from the message [Bri78].

P *A message containing a contrast to its intended setting will be more likely attended to by the user than a message that is harmonious with its setting [Bri78].*

P *A user will more likely pay attention to a message if he has few distraction in his environment than if his attention is divided.*

7.2. Correspondence with Graphics

7.2.1. Sound and Text

It was stated in Chapter 2 that earcons can either reinforce or replace their graphical counterparts. Certain messages are most effective if they are represented using both sight and sound. The simultaneous presentation of sound with graphics can strengthen the best characteristics of each of them. It is these types of messages that the following guidelines concern.

P *The sound messages present in an interface must reflect the nature of the tasks that the computer currently displays.*

P *An audio cue's affective image and the meaning of its corresponding textual message should relate.*

For maximum understanding, the senses of vision and hearing must reinforce each other. They must never be allowed to conflict.

P *If an audio message is to reinforce a textual message, then there must be precise timing of their simultaneous presentation.*

7.2.2. Sound and Color

Colors have a symbolic meaning and a psychological effect upon their observers. It has been shown [Bri78] that a significant relationship exists between the colors used in a message and the meanings that become attached to it. These relationships can be of great use in a sight-sound interface. If the appropriate colored text and audio cue are used in conjunction to express a particular message, then the computer user can more easily assume, learn, and remember the correct meaning of that message.

Studies have shown [Bri78] that the use of color can be a determining factor in learning and remembering. Color adds realism and carries moods and symbolism. They are capable of capturing attention, arousing feelings, and producing emotional responses. Color can be used to emphasize what needs to be emphasized.

In the Western culture, various associations have been determined for various colors. We will see in Chapter 10 that various associations also exist for various tones. Table 2 below integrate these two types of relationships to produce guidelines that specify the combined use of color and sound.

Graphics Color	Color Associations	Sound
red	danger, warnings, fire, excitement, anger	fast tempos
blue	calm, peace, official identifications	slow tempos
green	nature, fertility, freshness	high tones (middle C)
black	death, mourning, evil, formality	low pitches, minor key compositions
brown	masculinity, earth	low tones
white	purity, health, medicine, cleanliness	simple melodies
yellow	joy, happiness, sunshine, daylight	high pitches
pink	femininity	lively, high notes
purple	royalty	low pitches
violet/ blue violet	sadness	low pitched tones

Table 2

7.3. Correspondence with Computer Tasks

A user derives the meaning of an audio message partly from the nature of the computer event or task associated with it. Therefore, the types of tasks that a user is typically involved in must be analyzed. An interactive session with the user interface can be divided into three general areas:

(1). Dialogue Cycle

Borufka and Kuhlmann [Bor82] state that the interaction between a user and the computer can be built from the following four elements. Each element is an individual unit that completely describes a possible step in an interactive dialogue.

A. Prompt: specifies the type of data to be entered.

B. Input: input data.

C. Echo: shows interpretation of the entered input data.

D. Output: a returned value, error, or a request for more information.

(2). System Messages

This is supplementary information provided by the system to support an interactive session.

(3). Application Operations

These include standard operations on objects such as create, destroy, modify, view, select, search, store, and retrieve.

Each of the above categories of tasks should have a unique and identifiable type of audio cue associated with it. The user must be able to easily differentiate between the varying natures of the three categories of tasks. For example, the sound of a prompt should appear distinct from that of an echo. The user needs to be able to tell these apart as they each require different types of subsequent action. The audio cue associated with the task of a file being updated in a database should sound different than the sound of a file being retrieved from the database since modification and retrieval differ in their volatility. Earcons should be composed to fit both the situation in which it is to be placed and the task which it is to represent.

P *The greater the similarity between an audio message and it's associated computer task, the greater the degree and accuracy that a user perceives the correct meaning [Bri78].*

This is not to be interpreted to mean that earcons should always be represented by natural sounds. Many entities representable by earcons do not have a natural sound association. However, when such associations do exist, the above guideline is true.

P *Audio cues should represent a way of satisfying the current needs appropriate of the task which a user finds himself in when he hears the audio cue.*

P *The train of thought required for a particular computer task should not be broken if the interruption can be postponed [Eng75].*

A continuous train of thought is often required when a user is involved in tasks such as editing, debugging, searching, selecting, and problem solving. They should be able to carry out the task without constant interruption by the system. Audio cues can become a source of interruption if they are not composed of the sounds appropriate to the current task or if they are not timed and placed correctly within the sight/sound interface.

P *Tasks that require a steady steam of thought should mainly be supported by audio cues that are background and non-obtrusive in nature.*

If an interruption to such tasks is necessary, then and only then should loud audio messages be allowed to intervene.

PART V.

PRAGMATICS:

The Human Element of Earcons

CHAPTER 8

Learning and Remembering

8.1. Sound Organization

An earcon heard for the first time has no meaning and appears insignificant. A sequence of sounds only becomes significant as a result of the organizing capacities of the human brain [Dav78]. The ear's role in this process is as an intermediate agent that relays the sound from an external source to the brain. It does not organize them. The ear simply picks up the sound signal and thereafter the human mind constructs the meaningful earcon from the raw sound elements.

An earcon is not an audio cue simply by nature of its physical components. Earcons are a composition of interrelated tonal patterns and not a series of unrelated tones. An audio message is any tonal sequence that can be organized into meaningful units of several tones and which consequently does not sound like a random collection of isolated tones. It is distinguished from mere noise by the act of being perceived as a meaningful by the listener.

How do we organize sound into meaningful perceptual units? Gestalt psychology sheds light on this question. One basic belief of Gestalt psychology [Dav78] is that a person can hear complete units and patterns within a sequence of tones, even though the sequence is made up of individual sound elements. In other words, people have the ability to perceive general patterns in strings of separate tonal events, thereby organizing sound into perceptual units. This perceptual ability seeks to impose the best organization possible on the incoming sound elements. Therefore, a listener does not hear a series of individual sound elements, they hear whole phrases and become aware of particular patterns by grouping the tones into perceptual units. If they can do this successfully, the previously insignificant sequence of tones becomes a meaningful and informative earcon to the listener.

In order for an earcon to be of use to the listener, it must be learned and remembered. Learning, an active and complex process, is the partner of remembering. Remembering can be thought of as a continuation of the learning process. People can not be asked how well they learned something without assuming they remembered it. Learning and remembering represent the process by which listeners acquire and retain sound information that is to be associated with some external event.

Both learning and remembering decrease with the passage of time. Psychological studies indicate that people can forget as much as 60% within one hour after learning it [Wri66]. This can be reversed through the repeated exposure to the item to be learned and through distributed practice of learning the item. Therefore, the frequency of exposure during the training period is the key to learning and remembering an earcon.

P *Distributed repetitions of an earcon are more effective in causing a user to learn the earcon than massed repetitions of the earcon [Bri78].*

It should be noted that there is a limit to the effectiveness of frequency on learning. Overlearning, or overexposure, occurs when something is repeated and practiced a certain number of times over that necessary to learn it the first time. Degrees of overlearning beyond 50% (repeating 1/2 as many times as necessary) have been found inefficient as far as retention is concerned [Lun53].

P *Repetitions aid in the learning of an earcon, but repetitions of the same audio cue beyond a critical point causes overexposure resulting in no improvement in learning [Bri78].*

8.2. Sound Templates

Remembering is facilitated by the existence of a mental structure that is used to store information on audio events. Scientific evidence exists [Dav78] to suggest that an internal representation of the organized perceptual units composing an earcon resides in a listener's memory. The exact structure of this internal representation, or template, is coded during the learning process. A listener uses the internal representation to remember an earcon they already know and to recognize incoming earcons. When an incoming earcon is heard, it is compared to existing templates of known earcons. If the sequence of sounds is sufficiently similar to one of the existing templates, then it is recognized as an already learned earcon. If the sequence does not match any of the existing templates, then it is unrecognized and can be considered as a new sequence to the listener. In order for this new earcon to be remembered, it must go through the learning process. The ability to learn and remember an audio cue is related to the ability to produce an internal representation of the sequence against which incoming sounds can be compared.

The learning process of a newly encountered earcon is the process by which its unique internal template is created and stored in the users memory. Consider the situation where the user has heard an earcon and it is determined that it is unfamiliar. In other words, the incoming earcon does not match any of the existing templates and it is therefore not recognized nor remembered. After a series of encounters with this same earcon, its organizational structure is determined and coded into a unique internal representation. Then this newly created template is stored in memory. The new earcon has now been learned, it is fully remembered and recognizable.

The amount of time that the new template remains in memory is a function of many variables. One of these is the length of time elapsed between successive encounters of this same earcon. Memory has a purge time associated with it [Bri78]. If a remembered earcon is not encountered again during a certain amount of time, its template begins to deteriorate and slowly progresses to being forgotten.

The length of time it takes for an earcon to be processed from its audio reception to memory storage also varies. Earcons composed of shorter sequences of tones take less time to code and store. They therefore are processed faster than longer length audio cues.

P *Earcons composed of shorter sequences of sound elements are easier to remember than those composed of longer sequences of sound elements [Dav78].*

We have said the ability of the brain to group sound into its organizational units facilitates a listener's ability to remember sound information. Miller [Mil56] describes the process of memorization as the internal formation of chunks, or groups of related tones that go together, until there are few enough chunks so that we can recall all the tones in an earcon. An earcon designer can assist this internal organization ability by grouping related sound elements together. This way the incoming earcon is somewhat organized before it reaches the brain.

P *It is easier for a user to define a mental template for an audio cue if enough cues exist at the outset for them to be able to develop an internal representation* [*Bri78*].

The following guidelines are results of scientific experimentation that show what type of a priori external organization will assist the brain in its internal organization of sound information.

P *Earcons composed of related and organized groups of sound elements are easier to recognize and remember than those composed of unrelated sequences of tones* [*Dav78*].

M *Earcons organized into chunks of 5-6 unrelated tones or 10-12 related tones are easiest for the user to process and remember* [*Lun53*].

The above optimal numbers of sound elements for a group are the result of experiments done by Kwalwasser-Dykema and Drake [Lun53]. Miller

[Mil56] states a similar finding that a user can best remember sound information grouped into chunks of 5 to 9 tones. The results of the study by Miller did not state if the tones involved are related or unrelated. This could explain the difference between the two studies.

The ideas above can best be understood with an analogy in the English language. It is the relation between sense and nonsense verbal material. Letters presented as meaningful sentences or phrases are easiest to remember due to the relationships between the words. The next easiest to remember presentation is separate words. The most difficult way to remember letters is when they are presented as individual, unrelated letters. This same analogy holds for earcons. The easiest earcons to remember are those whose sound elements are related and grouped into meaningful chunks. The hardest earcons to remember are those whose entire composition consists of individual, unrelated tones.

M *Earcons composed of a simple repeated configuration involving only three different note values are easiest to remember [Lun53].*

M *Earcons composed of rhythms having four or more notes in varied patterns are of intermediate difficulty to remember [Lun53].*

P *Earcons composed of complicated rhythms and note relations are harder for the user to remember [Lun53].*

8.3. Space Complexity and Memory

Space complexity refers to the amount of space required for a user to remember an earcon. The goal is to minimize the space complexity for each audio cue in the computer/user interface. The less a user must remember, the easier it is for him to later recall and identify the learned audio cues.

Donald Knuth [Knu84] examines the complexity of songs and uses the words of songs as the units of his complexity measures. Before translating his results to sequences of notes rather than sequences of words, let us examine some of his song complexities.

A song of length n words without any type of repetitive structure has space complexity n. This complexity can be reduced by the use of a refrain. A refrain imposes an organization or relationship on the entire song thereby eliminating the amount of memory necessary to successfully remember a song. Knuth shows that when a song has a refrain, its space complexity can be reduced to $c*n$ where $c<1$. Songs with space complexity $O(\sqrt{n})$ are not uncommon and use the trick of repeating both refrains and verses. "Old MacDonald" has an even lower complexity accomplished by substituting the names of animals and animal sounds into similar verses as well as repeating the refrain and previous verses. "Partridge in a Pear Tree" has still better complexity of $O((\frac{n}{\log n})^{1/2})$ while "m Bottles of Beer on the Wall" has complexity of O(log n).

A song is sung to give listeners some kind of musical experience and is generally not used to convey information.[1] The space complexity of earcons is, perhaps, more important than of songs since earcons always convey

[1] Hymns and patriotic music are exceptions.

information. An earcon should convey the necessary information within as little user memory as possible, yet the result should not be unpleasant nor fatiguing.

P *An earcon should have low space complexity.*

> Sound messages having low space complexity take up a reduced amount of space in a user's memory and are easier for the user to learn and remember [Knu84].

Consider the space complexity of an earcon as the number of notes in that earcon. An earcon of length n notes, with no relationship or organizational structure present in the sequence of notes, has space complexity O(n).

This complexity can be reduced by the use of an organizational structure such as those described in the construction techniques of Family Earcons (section 5.3.4) and Compound Earcons (section 5.3.3). In a three motive family earcon, for example, the second motive is related to the first by a change in pitch and the third motive is related by a change in timbre. A user has less to remember since the second and third motives are only slight variations of the first motive. The space complexity of Family Earcons is the complexity of the family motive plus changes in pitch and timbre, which are of complexity 1.

Compound earcons greatly reduce the space complexity by using similar audio elements to represent similar features shared by a set of computer entities. A user has less to remember since a particular common feature is always represented by the same audio element. In the example given in section 5.3.3, a user would only have to remember the four audio cues for create, destroy, file, and string but could easily recognize and identify the four other earcons for create file, create string, destroy file, and destroy string.

CHAPTER 9

Levels of Experience

When designing a system using earcons in the sight/sound interface, the ideas of complexity level, information content, and user level must carefully be considered. The complexity level and the information content vary according to the user's level of experience [Dav78]. Complexity is a subjective quantity and must be measured from the observers point of view. What is complex to one person may not be so to another. It is dependent on the observer's experience with the object or subject in question.

A subject is usually considered complex by those who are unfamiliar with it. To the novice user, all audio cues or earcons appear highly complex. Novice users are inexperienced with the sight/sound interface. They haven't yet learned the earcons and their associated meanings. On the other hand, earcons are not very complex to the expert user of such a system. They are experienced with the system and have learned all the associated aspects of each earcon.

Complexity can also be described with respect to information content. Different amounts of information are present in an object because people perceive them as having varying degrees of subjective complexity [Dav78]. High information content is related to high complexity and low information content to low complexity. A highly complex object is one carrying new and useful information to its observer. As the observer becomes more experienced and acquires some of the information represented, this same object becomes less complex and carries less new information to that observer. An object with low complexity is one carrying little or no new information. As this object becomes more and more familiar, the information content decreases to the point where the observer begins to ignore the object completely.

An audio cue is of high complexity and high information content to the novice user. Every sound is important and carries new information. The novice must hear the entire earcon, possibly several repetitions, in order to recognize it. As they learn each earcon's sound, meaning, and associated computer event, this high complexity level begins to decrease. When it has decreased to a point of representing low information content, the novice is experienced and familiar with the earcons. They are no longer novices, but are now more advanced.

Likewise, to the expert, most earcons are of low complexity and low information content. Often times they don't need to hear the entire earcon to recall its tones, meaning, and associated computer event. They have learned it to the point of immediate mental recognition and recall after just a few notes. However, it should be noted that if the audio cue does not have a higher organization, or pattern, then even the expert may still need to hear the entire earcon in order to recognize and comprehend it.

We have stated that the complexity level and the information content level vary according to the level of the user. A terminal in this system must be adaptable to the current user. Different configurations are needed for different levels of users. A user must be able to configure his terminal to novice, intermediate, or expert level. In this way, the appropriate sight/sound interface will be in effect during their interactive sessions. As a user progresses from novice to expert, they can indicate their progress to the system and the system will subsequently upgrade to match the current level of experience.

The following guidelines concern the configuration of the system for different levels of user experience with respect to complexity and information content. Three levels are distinguished; novice, intermediate, and expert. These occur on a continuum with the largest percentage of users being in the intermediate range.

9.1. Novice

Novice users are inexperienced and unfamiliar with a sight/sound interface. They are just starting the learning process and currently do not recognize the audio cues and do not know their intended meanings. All earcons are of high information content and of high complexity to the novice user.

M *Earcons should be played at slower speeds to give the novice time to fully hear, learn, and recognize the earcon.*

M *Earcons may need more repetitions to give the novice user time to hear, learn, and recognize the earcon.*

M *Earcons for the novice user should be reinforced with a corresponding visual icon or textual message.*

M *Very short audio cues may need more repetitions than longer earcons.*

Due to half listening, a novice may possibly miss the entire audio cue if it is of short length. Therefore it may need to repeated. This may not be the case for more advanced users since a known audio cue can be recognized within a few notes of its hearing.

9.2. Intermediate

This group contains the largest percentage of the user population. Intermediate users are able to recognize and understand some but not all of the earcons in the system. They feel fairly comfortable with the system and understand the major purpose and usefulness of a sight/sound interface.

All audio cues are of intermediate complexity and represent an intermediate amount of information to this level of user. Studies have shown [Vit66,Wal73] that, on the whole, people prefer tunes that provide them with an average amount of information and consequently reduce their uncertainty about the associated events. They do not like ones which, either because of insufficient or too much information, do not reduce their uncertainty about events. The graph in figure 11 shows Vitz's [Vit72] hypothesized relationship between preference and complexity level.

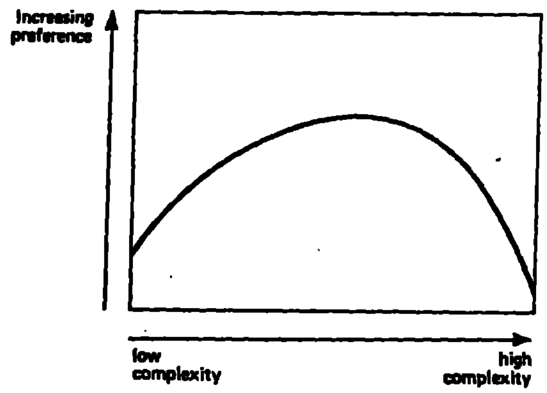


Fig. 11

The intermediate group of users represents the "average" or intermediate amount of information, neither too much nor too little. This is the target group (average user) that most earcons are designed for. The guidelines presented in Chapters 4, 5, 6, 7, and 8 were aimed at the intermediate level user.

9.3. Expert

Expert users are very experienced with a sound interface. They know each earcon's sound, meaning, and associated computer event. They are able to recognize an earcon without hearing its entire length. All earcons are of low complexity and of low information content to the expert user.

M *Often used earcons may be played at faster speeds for the expert user since they can recognize and recall them only within a few tones.*

M *Audio cues may not need as many repetitions as required for the novice user due to an expert's immediate recall ability.*

CHAPTER 10

Effect on User

10.1. Psychological Responses

A computer user learns the intended meaning of an earcon through a process of conditioning. Since earcons are nonverbal audio events, the meaning is manifested in the way a particular earcon affects the listener emotionally. Through certain learned associations with particular events, as earcon can come to evoke emotional feelings and responses.

Davies [Dav78] states that almost any piece of music can become associated with any emotional event, regardless of the nature of the event. People learn through training and experience that certain types of sounds go with particular types of events. This lends evidence to the fact that, after a period of learning and experience, computer users can become accustomed to the presence, use, and meaning of audio cues in a computer-user interface.

The emotions felt by the listeners are not intrinsic to the sound but they are a property of the listener. The emotional responses become attached to the sound through training. The fact that it is learned, however, does not make it any less significant or real.

The associations a listener does learn between a sound and a particular event is culturally dependent. Sound which does not use conventions from our Western culture are unfamiliar and thus might sound strange and confusing to Western ears. However, it should not be assumed that they are meaningless and strange in the absolute sense. The reason it appears that way is because we are unfamiliar with sounds composed using non-western conventions.

Consequently, the user of a computer system employing the use of earcons may at first feel uncomfortable with the system and think that the sounds are nonsense. However, after a period of learning which earcons are associated with which computer events, the earcons will begin to assume familiarity and meaningfulness. What a user found insignificant at one time will begin to acquire the more useful properties of earcons and the users liking of the system will consequently increase with repeated hearing.

Table 3 was compiled from results [Lun53] obtained by experiments done on subjects listening to various conventions of Western music. It was reported that the subjects had no difficulty in describing their emotional reactions to the sound stimuli. One significant aspect of all the reported findings is the striking similarity among responses reported by the large majority of subjects regardless of their age, intelligence, or prior training. These findings have been generalized to show how they can be used to compose earcons that produce fairly predictable user emotional feelings and/or responses.

Tonal Characteristics	Affective Responses
major key compositions	bright, cheerful, joyful, exuberant
minor key compositions	gloomy, plaintive, melancholic, mournful
slow tempos	dignified, calm, serene, sentimental, tender
fast tempos	happy, gay, exciting, restless
high pitch compositions	lively, humorous
low pitch compositions	vigorous, majestic, dignified, serious, sad
rhythm with strong beats	spiritual, lofty
smooth flowing rhythms	light, happy, playful
complex, dissonant harmonies	exciting, agitating, vigorous, inclined toward sadness
simple, consonant harmonies	happiness, graceful, serene, lyrical

Table 3

For example, earcons eliciting dreamy-sentimental reactions could be composed using minor modes, slow tempos, flowing rhythms, and sometimes simple harmonies and high pitches [Lun53]. Earcons eliciting happiness requires fast tempos, simple harmonies, and flowing rhythms [Lun53].

10.2. Physiological Responses

The sounds of music have surprising physiological affects on the human body. Studies [Lun53] have found that music tends to cause changes in breathing, blood pressure, and blood supply. Rhythmic and strongly vigorous sounds tend to have a greater affect on the increase of these physiological factors than other types of compositions. This type of music favorably influence the cardiovascular system, digestion, working power, and other body functions.

Although these "musical affects" are precisely the sort of reactions and associations to avoid by keeping the earcons short and as melodically neutral as possible, it is important that interface developers be aware of the possible existence of such factors because their goal is the produce safe and effective computer-user interfaces. Through knowledge of the psychological and physiological affects of music and sound in general, any negative affects can be avoided.

P *Loud sequences of tones have a greater effect on accelerating the breathing rate than soft tones [Lun53].*

P *The more definite the melody and rhythm in a tonal composition, the greater the chances that physiological reactions will occur [Lun53].*

P *Compositions consisting of "marches" create the greatest rise in pulse and blood pressure [Lun53].*

PART VI.

EPILOGUE

CHAPTER 11

Concluding Remarks

As we have seen, sound has many useful aspects allowing it to play a very powerful role in computer systems. The integration of both sight and sound in computer/user interfaces provides the user with a natural communication environment yielding effective and efficient man-machine interactive sessions.

This thesis presents an extensive look into the world of sound and introduces a new approach of incorporating sound into the computer/user interface. To our knowledge, this kind of study has not been done before. A new concept, "earcons", was born out of a synthesis of information found in various disciplines such as acoustics, computer science, psychology, linguistics, communication, and music. One possible construction technique for earcons is discussed in detail although other techniques are just as plausible. Perhaps future research will yield proven techniques and standards in this area. Various human issues of sound messages are also examined.

The next step in the continuation of this research is the development of an actual implementation of a computer/user interface employing earcons. Once this is done, human factors experimentation can be performed on its users to substantiate and refine many of the suggestions and guidelines proposed in Chapters 5, 6, 9, and 10. It should be noted that many of the guidelines can be individually human factors tested before any specific implementation is developed. For example, a music synthesizer can be used to create specific earcons, which are taped and played back under test conditions.

This helps to achieve a "good" prototype by alleviating certain design problems beforehand.

Due to the size limitation of a Masters Thesis, not all aspects of earcons could be fully investigated or even addressed. For the sake of future research, however, a few are worth mentioning. The use of voice has many interesting possibilities in a sight and sound user interface. Earcons could be composed entirely of synthesized voice or non-voice earcons could be used to reinforce voice output and/or graphics. Advertising often uses combinations of voice and nonverbal sounds. Also, voice recognition as sound input could result in complete two-way man-machine communication. Earcons for different types of systems could also be investigated. For example, would the structure and composition of earcons for office use be different than earcons for scientific use? The details of physical implementation of earcons in a user interface is an area in itself and requires substantial research. Since earcons are a communication language utilizing nonverbal sound messages, a formal grammer could be defined for their composition. Formal grammers have been successfully established for music [Roa79].

Research needs to be done into the differences between icons and earcons. This study suggests the possibility of using earcons to replace icons but there are also situations where an iconic equivalent does not exist for an earcon application. Since this is a project in itself, it was not incorporated into the thesis.

The utilization of sound in computer systems is still in its infancy but its life expectancy is promising. Many contemporary systems already provide some type of sound producing hardware and make use of sound in a minimal way. As a secondary sensory stimulus, PICT [Gli84] currently employs single audio cues in its user interface. It uses a single sounding of a bell to confirm acceptance of user commands and two soundings of a tone in rapid succession to draw attention to an error. Human factors testing of PICT on more than sixty subjects found that 78% of its users thought sound was important and useful in interactive computer sessions.

One purpose of the study is to provide a basic set of design guidelines and to present some initial ideas that will stimulate interest and motivate further research in this area. This thesis attempts to realize this goal by placing a solid stepping stone into the evolutionary path of sight and sound computer/user interfaces.

Bibliography

[Ber66] Bernstein, Martin and Martin Picker. *An Introduction to Music*. 3rd ed., Englewood Cliffs: Prentice-Hall, 1966.

[Bla76] Blake, Ian and Ronald C. Mullin. *An Introduction to Algebraic and Combinatorial Coding Theory*. New York: Academic Press, 1976.

[Bor82] Borufka, H.G. and H.W. Kuhlmann. "Dialogue Cells: A Method for Defining Interactions". *IEEE*, July 1982, 25-33.

[Bri78] Britt, Steuart Henderson. *Psychological Principles of Marketing and Consumer Behavior*. Lexington: Lexington Books, 1978.

[Bro54] Broadbent, D. E. "*Effects of Noise of High and Low Pitch on Behavior*". Report No. APU 222, Cambridge: Medical Research Council, Applied Psychology Unit, 1954.

[Com82] *Commodore 64 Programmer's Reference Guide*. Indianapolis: Howard W. Sams & Co., Inc., 1982.

[Chr66] Christ, William, Richard DeLone, Vernon Kliewer, Lewis Rowell, and William Thomson. *Materials and Structure of Music I*. Englewood Cliffs: Prentice-Hall, Inc., 1966.

[Dav78] Davies, John Booth. *The Psychology of Music*. Stanford: Stanford University Press, 1978.

[Deu82] Deutsch, Diana. *The Psychology of Music*. New York: Academic Press, 1982.

[Eas70] Easterby, R.S. "Perception of Symbols for Machine Displays". *Ergonomics*. 1970, 13, 149-158.

[Eng75] Engel, Stephan E. and R.E. Granda. *Guidelines for Man/Display Interfaces*. IBM Technical Report TR 00.2720. Poughkeepsie: IBM Poughkeepsie Laboratory, 1975

[Gar53] Garner, W.R. "An Informational Analysis of Absolute Judgements of Loudness". *Journal of Experimental Psychology*., 1953, 46, 373-380.

[Gel53] Geldard, Frank A. *The Human Senses*. New York: John Wiley & Sons, Inc., 1953.

[Gli84] Glinert, Ephraim P. and Steven L. Tanimoto. "PICT: An Interactive Graphical Programming Environment". *Computer*, November 1984, 17, 11, 7-25.

[Gre85] Greenberg, Robert. University of California at Berkeley, private communications, 1984-1985.

[Har34] Hartshorne, Charles. *The Philosophy and Psychology of Sensation*. Chicago: The University of Chicago Press, 1934.

[Hei83] Heilborn, John and Ran Talbott. *Your Commodore 64: A Guide to the Commodore 64 Computer*. Berkeley: Osborne/McGraw-Hill, 1983.

[Hem82] Hemenway, Kathleen. "Psychological Issues in the Use of Icons in Command Menus". *Proceedings of Human Factors Society 26th Annual Meeting*, Seattle, WA., October 1982.

[Hug74] Huggins, W.H. and Doris R. Entwisle. *Iconic Communication: an annotated bibliography*. Baltimore: The Johns Hopkins University Press, 1974.

[Ker80] Kerman, Joseph. *Listen*. New York: Worth Publishers, Inc., 1980.

[Knu84] Knuth, Donald E. "The Complexity of Songs". *Communications of the ACM*, 27 (April 1984): 344-348.

[Kol69] Kolers, Paul. "Some Formal Characteristics of Pictograms". *American Scientist*, 57, 3, 348-363, 1969.

[Lun53] Lundin, Robert W. *An Objective Psychology of Music*. New York: The Ronald Press Company, 1953.

[Man84] Mansur, Douglas. "*Graphs in Sound: A numerical Data Analysis Method for the Blind*". Masters Thesis, University of California, Davis, 1984.

[Mar84] Marcus, Aaron. "Corporate Identity for Iconic Interface Design: The Graphic Design Perspective". *Computer Graphics and Applications*, December 1984, 4, 12, 24-32.

[Mat69] Mathews, Max V. *The Technology of Computer Music*. Cambridge: The M.I.T. Press, 1969.

[Mey56] Meyer, Leonard B. *Emotion and Meaning in Music*. Chicago: The University of Chicago Press., 1956.

[Mil56] Miller, George. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information". *The*

Psychology Review. March 1956, 63, 2, 81-97.

[Mye84] Myers, Brad A. "The User Interface for Sapphire". *Computer Graphics and Applications*, December 1984, 4, 12, 13-23.

[Pie56] Pierce, John R. *Man's World of Sound*. New York: Doubleday & Company, Inc., 1956.

[Plo65] Plomb, R. and W.J.M. Levelt. "Tonal Consonance and Critical Bandwidth". *Journal of the Acoustical Society of America*, 1965, 38, 548-560.

[Pol52] Pollack, I. "The Information of Elementary Auditory Displays". *Journal of Acoustical Society of America*, 1952, 24, 745-749.

[Pol53] Pollack, I. "The Information of Elementary Auditory Displays II". *Journal of Acoustical Society of America*, 1953, 25, 765-769.

[Ram79] Ramsey, H. Rudy and Michael E. Atwood. *Human Factors in Computer Science: A Review of the Literature*. Technical Report SAI-79-111-DEN. Englewood, CO: Science Applications, Inc., 1979.

[Ras82] Rasch, R.A. and R. Plomb. "The Perception of Musical Tones". in Deutsch, Diana (ed.), *The Psychology of Music*. New York: Academic Press, 1982.

[Rat66] Ratner, Leonard G. *Music: The Listener's Art*. New York: McGraw-Hill Book Company, 1966.

[Roa79] Roads, Curtis. "Grammers as Representations for Music". *Computer Music Journal*, 1979, 3, 1, 48-55.

[Ros76] Ross, Sheldon. *A First Course in Probability*. New York: Macmillan Publishing Co., Inc., 1976.

[Ros82] Rossing, Thomas D. *The Science of Sound*. London: Addison-Wesley Publishing Co., 1982.

[Rya56] Ryan, T.A. and C.B. Schwartz. "Speed of Perception as a Function of Mode of Representation". *American Journal of Psychology*, 1956, 69, 60-69.

[Sch76] Schiffman, Harvey Richard. *Sensation and Perception: An Integrated Approach*. New York: John Wiley & Sons, Inc., 1976.

[Sin83] Singh, Baldev, John C. Beatty, and Rhonda Ryman. "A Graphics Editor for Benesh Movement Notation". *Computer Graphics*, 17, 3 (July 1983), 51-62.

[Smi82] Smith, David Canfield, Charles Irby, Ralph Kimball, Bill Verplank, and Erik Harslem. "Designing the Star User Interface". *Byte: The Small Systems Journal*, April 1982, 242-282.

[Til76] Tilbrook, D. *A Newspaper Page Layout System.* M.Sc. Thesis, Department of Computer Science, University of Toronto, Toronto, Ontario, 1976.

[Vit66] Vitz, P.C. "Affect as a Function of Stimulus Variation". *Journal of Experimental Psychology*, 1966, 71, 1, 74-79.

[Vit72] Vitz, P.C. "Preference for Tones as a Function of Frequency (Hertz) and Intensity (Decibels)". *Perception and Psychophysics*, 1972, 2 (1B), 84-88.

[von66] von Fieandt, Kai. *The World of Perception.* Homewood: The Dorsey Press, 1966

[Wal73] Walker, E.L. "Psychological Complexity and Preference: A Hedgehog Theory of Behavior". in Berlyne, D.E. and K.B. Madsen (eds.), *Pleasure, Reward, and Preference*. New York: Academic Press, 1973.

[Web77] *Webster's New Collegiate Dictionary.* Springfield: G. & C. Merriam Company, 1977.

[Wil84] Williams, Gregg. "The Apple MacIntosh Computer". *Byte: The Small Systems Journal*, February 1984, 9, 2, 20-40.

[Wil83] Williams, Gregg. "The Lisa Computer System". *Byte: The Small Systems Journal*, February 1983, 8, 2, 76-90.

[Wri66] Wright, John S. and Daniel S. Warner. *Advertising*. New York: McGraw Hill Inc., 1966

[Yeu80] Yeung, E.S. "Pattern Recognition by Audio Representation". *Analytic Chemistry*, 1980, 52, 7, 1120.